ERIC Notebook

September 1999 Issue 5

Case-control studies

UNC SCHOOL OF PUBLIC HEALTH DEPARTMENT OF EPIDEMIOLOGY

Michel Ibrahim, MD, PhD
Director of Education Program

Lorraine Alexander, DrPH

Carl Shy, MD, DrPH

Sherry Farr, Graduate Research Assistant

ERIC EPIDEMIOLOGIC RESEARCH & INFORMATION CENTER Durham, North Carolina

Ron Horner, PhD
ERIC Director

The ERIC Notebook is funded by the Department of Veterans Affairs (DVA), Veterans Health Administration (VHA), Cooperative Studies Program (CSP), to promote the strategic growth of the epidemiologic capacity of the DVA.

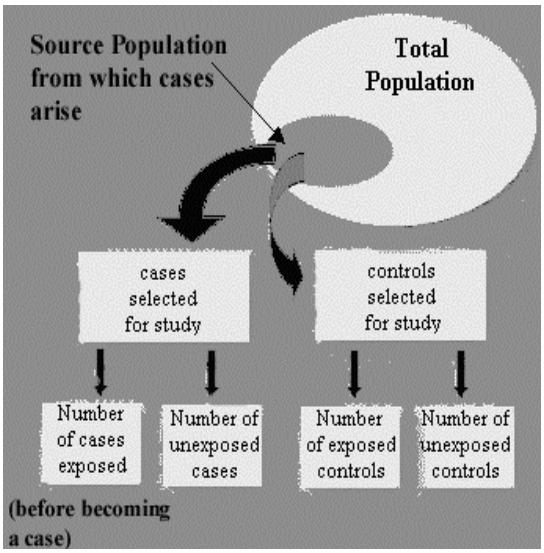
Case-control studies are used to determine if there is an association between an exposure and a specific disease. These studies proceed from effect (disease) to cause (exposure). Case control studies assess whether exposure is disproportionately distributed between the cases and controls, which may indicate that the exposure is a risk factor for the disease under study. Case-control studies are frequently used for studying rare diseases.

Unlike cohort or cross-sectional studies, subjects in case-control studies are selected because they have the disease of interest (cases). Selection is not based on exposure status. Controls, persons who are free of the disease under study, are randomly selected from the population out of which the cases arose.

At baseline:

- **Selection of cases and controls based on disease status.**
- **Exposure status is unknown.**

After cases and controls have been identified, the investigator determines the proportion of cases and the proportion of controls that have been exposed to the exposure of interest.



After the investigator determines the exposure table can be formed from the study data.

	Cases	Controls
Exposed	a	c
Unexposed	b	d

Measures of incidence in case-control studies

In case-control studies the proportion of cases in the entire population-at-risk is unknown, therefore one cannot measure incidence of disease. The controls are representative of the population-at-risk, but are only a sample of that population, therefore the denominator for a risk measure, the population-at-risk, is unknown. However, one can obtain a valid estimate of the relative risk (RR) by using the exposure odds ratio (OR).

- Odds of exposure among cases = a / c
- Odds of exposure among controls = c / d

Therefore, the exposure odds ratio is:

- $OR = (a \times d) / (b \times c) \approx RR$

Cohort study

Diseased person-years

	Disease	No disease
Exposed	a	n_1
Unexposed	c	n_2

$RR = (a / n_1) / (c / n_2)$

Case Control Study

	Cases	Controls
Exposed	a	c
Unexposed	b	d

$OR = (a/c)/(b/d) = (a/b)/(c/d)$

If b and d (from the case-control study) are sampled from the source population, $n_1 + n_2$, then c will represent the n_1 component of the cohort and d will represent the n_2 component, and $(a/n_1)/(c/n_2)$ will be estimated by (a/b)/(c/d).

Interpreting the Odds Ratio
The odds ratio is interpreted the same way as the relative risk that it approximates.

OR = 1 Disease risk is the same for exposed and unexposed
OR > 1 Exposure increases disease risk
OR < 1 Exposure reduces disease risk

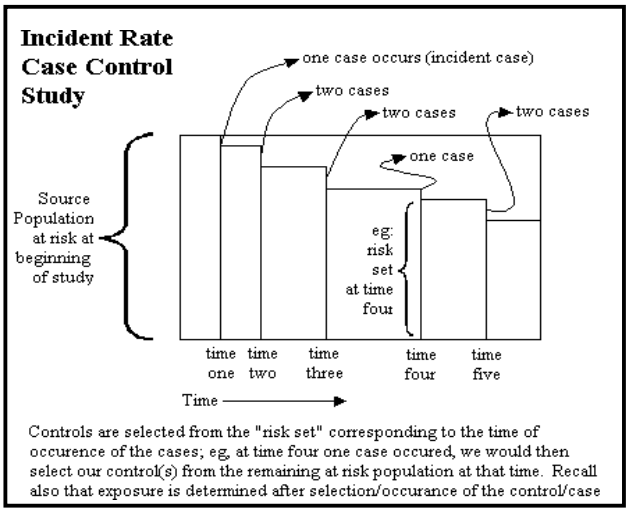
- For example, investigators conducted a case control study to determine if there is an association between colon cancer and a high fat diet. Cases were all confirmed colon cancer cases in North Carolina in 1997. Controls were a sample of North Carolina residents without colon cancer. The odds ratio was 4.0. This odds ratio tells us that individuals who consumes a high fat diet have four times the risk of colon cancer than do individuals who do not consume a low fat diet. In another study of colon cancer and coffee consumption, the OR was 0.60. Thus, the risk of colon cancer among coffee drinkers is only 0.60 times the risk among individuals who do not consume coffee. This OR tells us that coffee consumption is protective against colon cancer.

Types of case-control studies. Case-control studies can be categorized into two different groups based on when the cases develop disease. Some case-control studies use prevalent cases while other case-control studies use incident cases.

Prevalent cases are all persons with the disease during the observation period. For example, all cases of arthritis examined at a hospital between 1993-97. These studies yield a prevalence odds ratio, which will be influenced by the incidence rate and survival or migration out of the prevalence pool of cases, and thus does not estimate the rate ratio.

Cumulative incidence (CI) case-control studies require selection of incident cases. Controls are selected from individuals in the source population who never became a case during the entire period of observation. The odds ratio in these studies estimates the risk ratio of cohort studies if the proportion of diseased among each exposure group is less than 10% (requires rare disease assumption).

In incidence rate case-control studies selected cases are incident cases. Controls must be selected from the at-risk source population at the same time as cases occur and must be eligible to become a case if the disease develops in the control at a later time during the period of observation. In incidence rate case-control studies the odds ratio estimates the incidence rate ratio of cohort studies, without assuming that the disease is rare in the source population.



Note that it is possible, albeit rare, that a control selected at time two, for example, could become a case during the remaining time that the study is running. This differs from cumulative incidence case control studies, which select their controls after the conclusion of the study from among those individuals remaining at risk.

Selecting controls in this "risk set" or incidence rate manner provides two advantages:

1. A direct estimate of the rate ratio is possible

2. The estimates are not biased by differential loss to follow up among the exposed vs. unexposed controls.

- For example, referring back to the graphic above, if a large number of smokers left the source population after time four they would not be available for selection at the end of the study -- when CI controls would be selected. This would give the investigators biased information regarding the level of exposure among the controls over the course of the study.

Source populations for case-control studies. Source populations can be restricted to a population of particular interest, e.g. postmenopausal women at risk of breast cancer. This restriction makes it easier to control for extraneous confounders in the population. Controls should represent the restricted source population from which cases arise, not all non-cases in the total population. The cases in the study do not have to include all cases in the total population.

Sources of cases.

- Cases diagnosed in a hospital or clinic
- Cases entered into a disease registry, e.g. cancer, birth defects, deaths
- Cases identified through mass screening, e.g.

hypertensives, diabetics
- Cases identified through a prior cohort study, e.g. lung cancers in an occupational asbestos cohort

Sources of controls.
- Population controls are non-cases sampled from the source population giving rise to cases.
  This is the most desirable method for selecting controls.
- Neighborhood or friend controls are appropriate for selection as controls if these individuals would be included as cases if they developed the disease of interest. It is not appropriate to select neighbors or friends as controls if they share the exposure of interest.
- Hospital controls
  There are certain problems with hospital controls in that they may not be from the same source population from which the cases arose. Hospital controls may not be representative of the exposure prevalence in the source population of cases, e.g. there may be a higher prevalence of smokers in hospitals. Hospital controls also may have diseases resulting from the exposure of interest, e.g. the exposure (smoking) is related to the disease of interest (cancer) and to heart and lung diseases from which the controls may be suffering.
- Controls with another disease
  However if the study is on lung cancer it is essential to exclude cancers known or suspected to be related to the study exposure of interest. These controls also share some of the same problems as hospital controls.

Advantages of case-control studies. Case-control studies are the most efficient design for rare diseases and require a much smaller study population than cohort studies. Additionally, investigators can avoid the logistical challenges of following a large population over time. Thus, case-control studies also allow more intensive evaluation of exposures of cases and controls. Incidence rate case-control studies also yield a valid estimate of the incidence rate ratio derived from a cohort study if incident cases are studied and controls are sampled from the risk set of the source population.

Disadvantages of case-control studies. Case-control studies do not yield an estimate of the incidence rate or cumulative incidence, as the denominator of these measures is not defined. Case-control studies may also be subject to recall bias if exposure is measured by interviews and if recall of exposure differs between cases and controls. However, investigators can avoid this problem if historical records are available to assess exposure. Choosing an appropriate source population is also difficult and may contribute to selection bias. Case-control studies are not an efficient means for studying rare exposures (less than 10% of controls are exposed) because very large numbers of cases and controls are needed to detect the effects of rare exposures.

Self-evaluation

Q1: Suppose that in a case-control study using incident cases of colon cancer you found that 80% of the cases were married. Does this demonstrate that being married increases the risk of developing colon cancer?

Q2: In the same case-control study above, assume that 90% of the control group are married. If there are 200 cases and 200 controls estimate the relative risk of colon cancer for single men. Construct a 2x2 table and determine and interpret the exposure odds ratio.

Answers

1. No. In order to assess whether or not the exposure of interest (marriage) increases the risk of having colon cancer, the proportion of controls that are married must also be known and the exposure OR must be computed.

2.

Marital status	cases	controls	total
single	40	20	60
married	160	180	340
total	200	200	400

OR = $(40 \times 180) / (20 \times 160) = 2.25$

Glossary

Cohort Studies : An observational study in which subjects are sampled based on the presence (exposed) or absence (unexposed) of a risk factor of interest. These subjects are followed over time for the development of a disease outcome of interest. See ERIC Notebook issues 3 and 4 for details

Cross-Sectional Studies: An analytic investigation in which subjects are sampled at a fixed point for a period of time, and then the associations between the concurrent presence or absence of risk factors and diseases are investigated.

Exposure odds ratio (OR): the odds of a particular exposure among persons with a specific disease divided by the corresponding odds of exposure among persons without the disease of interest. Estimates the RR in case-control studies.

Incident case: a person who is newly diagnosed as a case.

Prevalent case: a person who has a disease of interest that was diagnosed in the past.

Relative risk (RR): the likelihood of a particular disease occurrence among persons exposed to a given risk factor divided by the corresponding likelihood among unexposed persons.

Source population: the population out of which the cases arose.

From: Medical Epidemiology, R.S. Greenberg, 1993, 1996.

If you would like to receive ERIC Notebook please fill out the form below:

Name: _____

Degree(s):_____

Address:_____

City, State, Zip:_____

Telephone Number:_____

Fax Number:_____

E-mail Address:_____

Please fax to: 919-966-2089 – Attn: Angie Tuck **Or**

Mail to : UNC-CH , School of Public Health , Attn: Angie Tuck Department of Epidemiology , Campus Box 7400, Chapel Hill, NC 27599

Upcoming Topics

Please let Angie Tuck know which topics are of special interest to you so that we can include them in a future issue.

UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
SCHOOL OF PUBLIC HEALTH
DEPARTMENT OF EPIDEMIOLOGY
CAMPUS BOX 7400
CHAPEL HILL, NC 27599